

Inferring Sequence Rules of Promoter Organization by Systematically Evaluating Biological Hypotheses

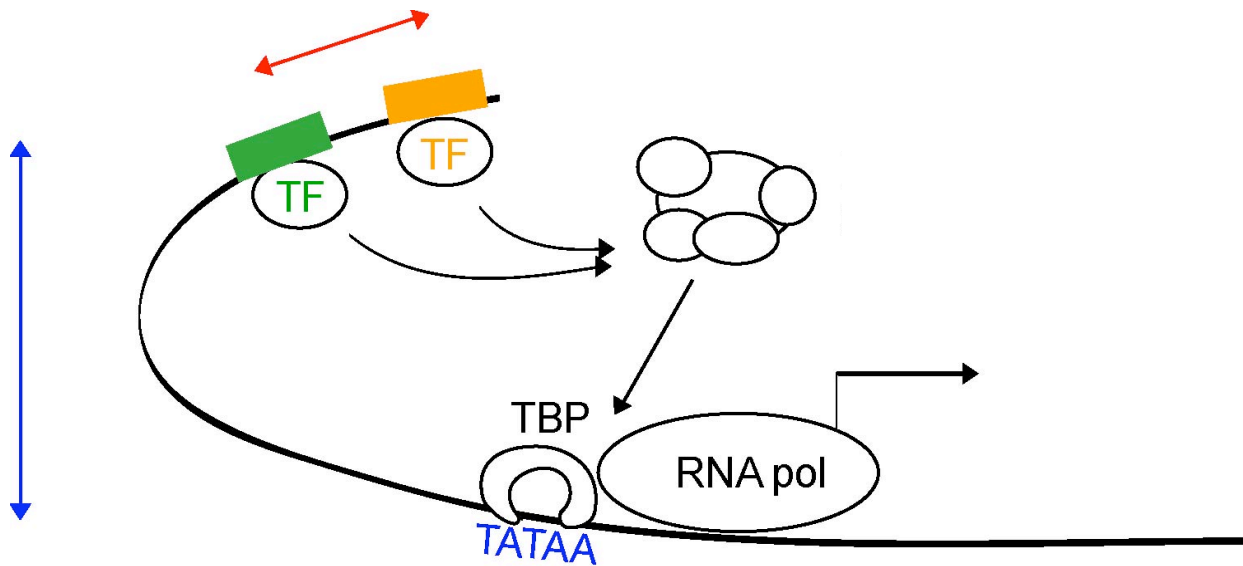
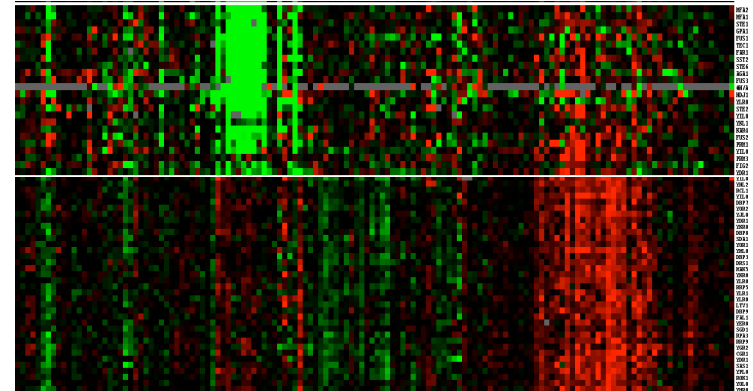
Derek Chiang

Michael Eisen Lab

University of California, Berkeley and
Lawrence Berkeley National Lab

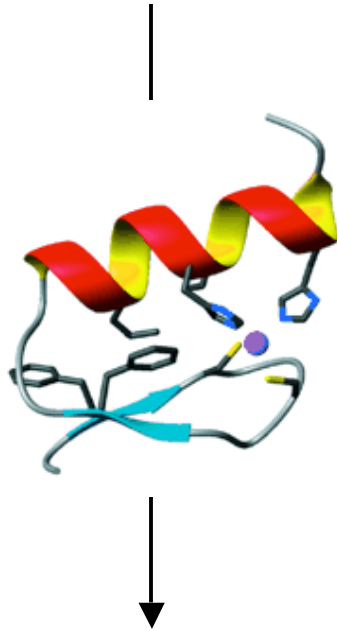
Promoter organization in yeast

```
>Saccharomyces cerevisiae chr V  
CCAATAGGTAAACTAAAATACAATAGAAGGG  
GTACTGAGTGCACGTGACTTATTTTTTTTTT  
TTGGTTTTAGGTTTCGCTTTTTTTCACCTTTT  
TCTACTTTCTAACACCACAGTTTTTGGGCGGG  
AAGCGGAAACGCCATAGTTGTAGGTCACCTGG  
CGTGAGTCAAGGCCGGGCAGCCAATGACTAA
```



From classification to mechanistic understanding

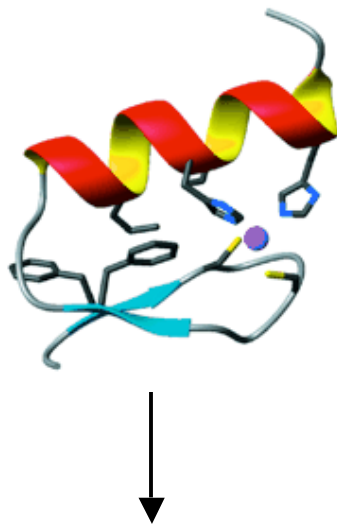
Fold recognition



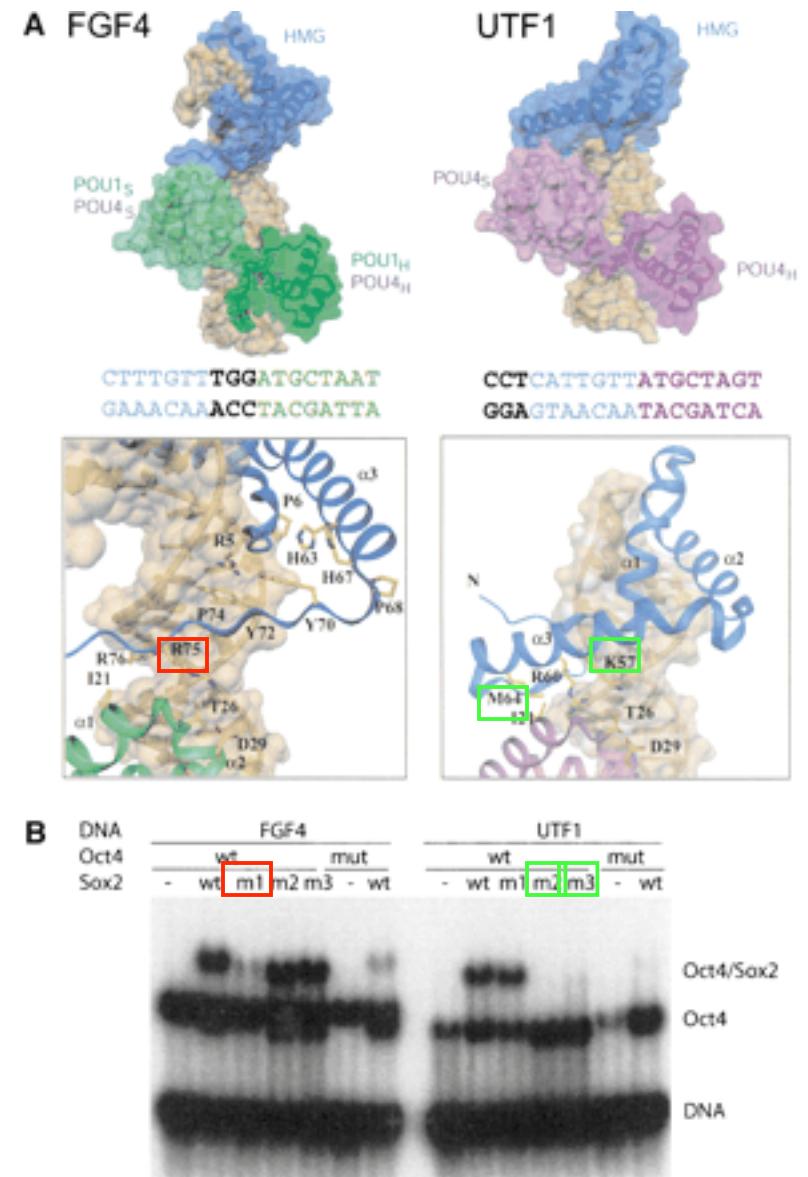
De novo protein design

From classification to mechanistic understanding

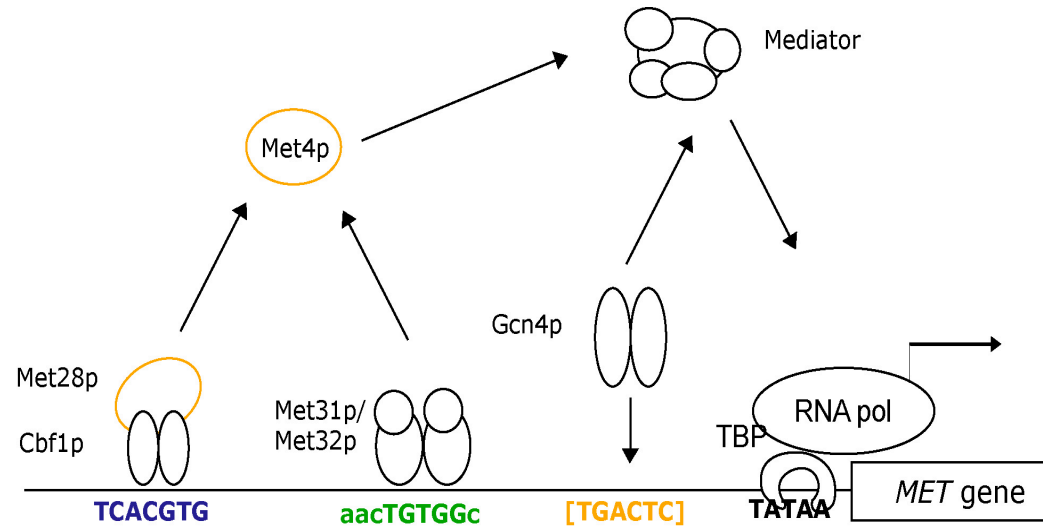
Fold recognition



De novo protein design



Case study: yeast sulfur utilization genes

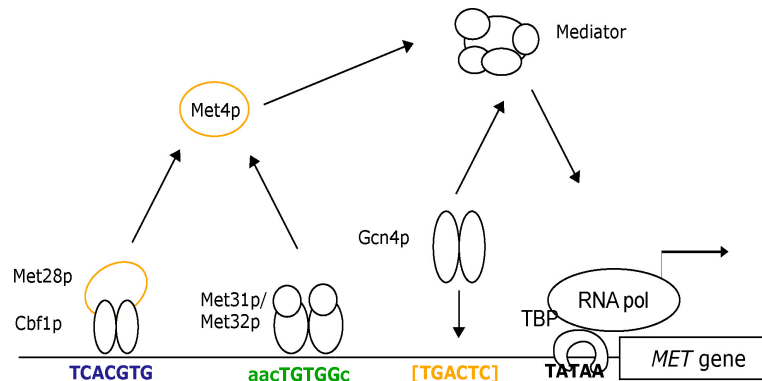


1) Genetic hypothesis testing

Predict Met4-regulated genes from microarray data

2) Define rules for promoter organization

Genetic hypothesis testing



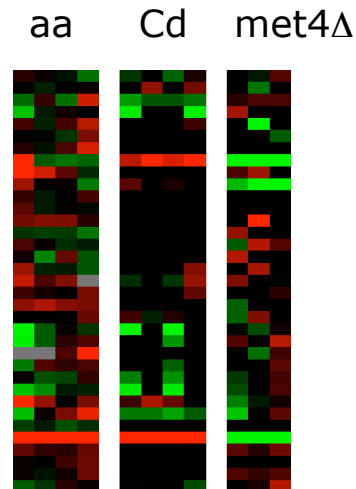
↑ Amino acid starvation

↑ Cadmium

↓ *met4*Δ in cadmium

Likelihood Ratio Test

$$\frac{\Pr(\text{Data} \mid \text{Model})}{\Pr(\text{Data} \mid \text{Background})} = \frac{\prod_g \prod_c p(E_{gc} \mid \mu_{m,c}, \sigma_c)}{\prod_g \prod_c p(E_{gc} \mid \mu_{b,c}, \sigma_c)}$$



vs.



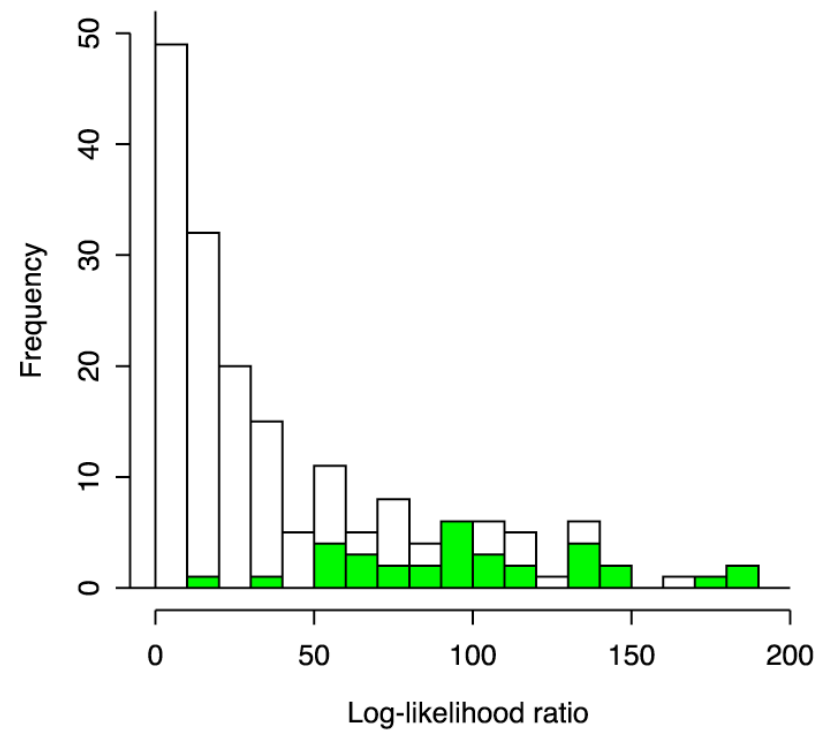
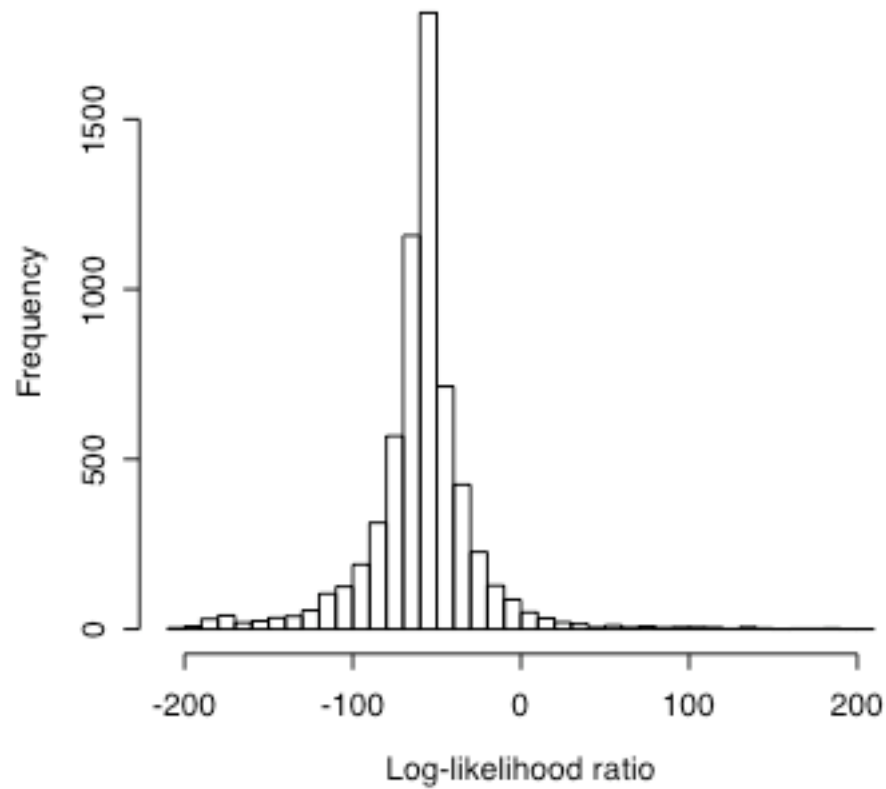
Model 1 ($\mu_{m,c}$)

Prototype for sulfur utilization gene

Background ($\mu_{b,c}$)

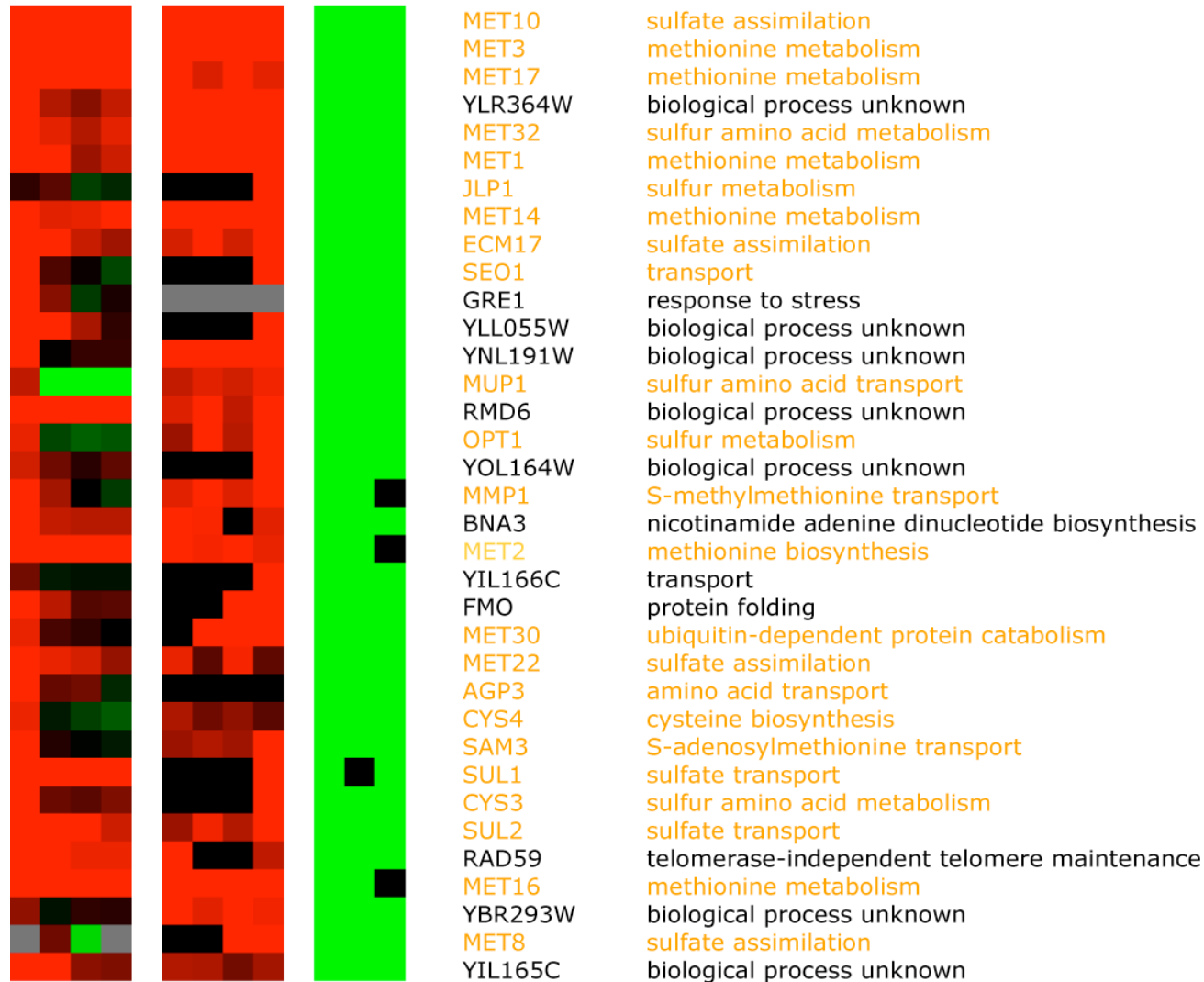
Genome-wide average

Genetic hypothesis testing accurately predicts sulfur genes



Genetic hypothesis testing accurately predicts sulfur genes

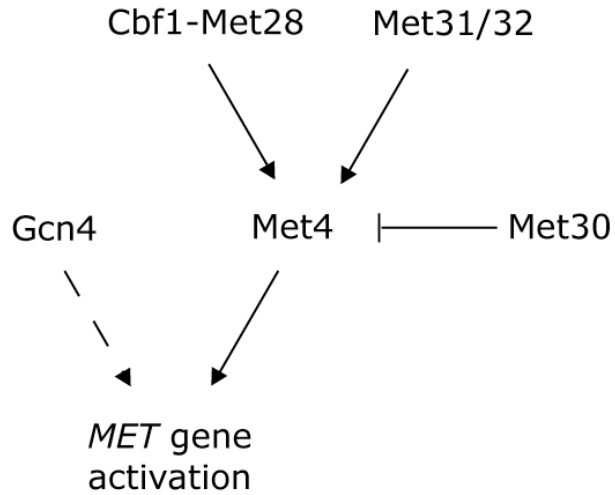
aa starv cadmium met4Δ



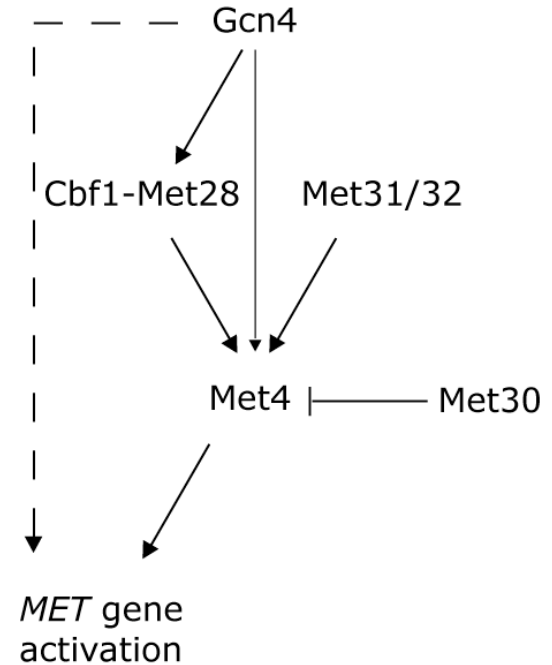
4x repressed 4x induced

Genetic models for Gcn4's role in regulation

Independent pathways

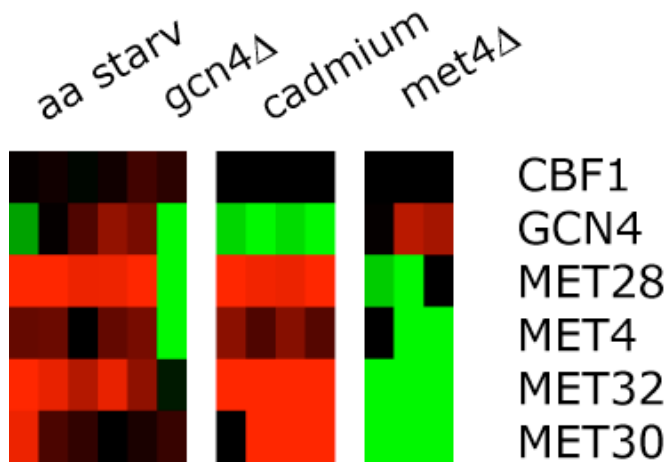
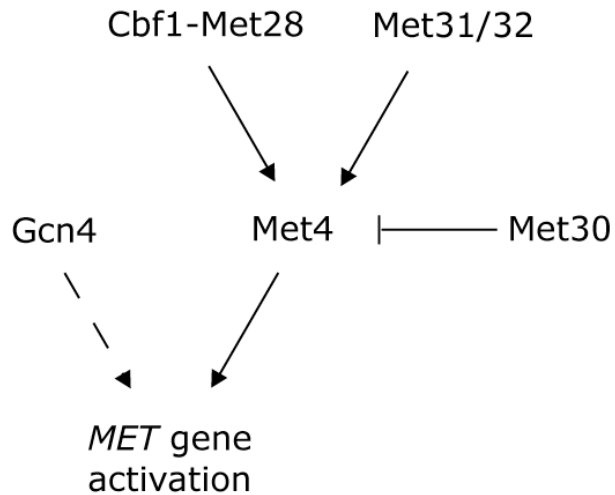


Feedforward (& feedback) regulation

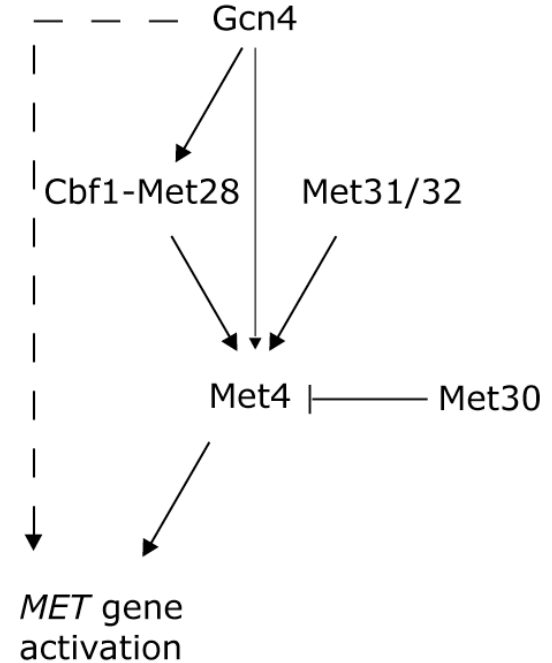


Genetic models for Gcn4's role in regulation

Independent pathways



Feedforward (& feedback) regulation



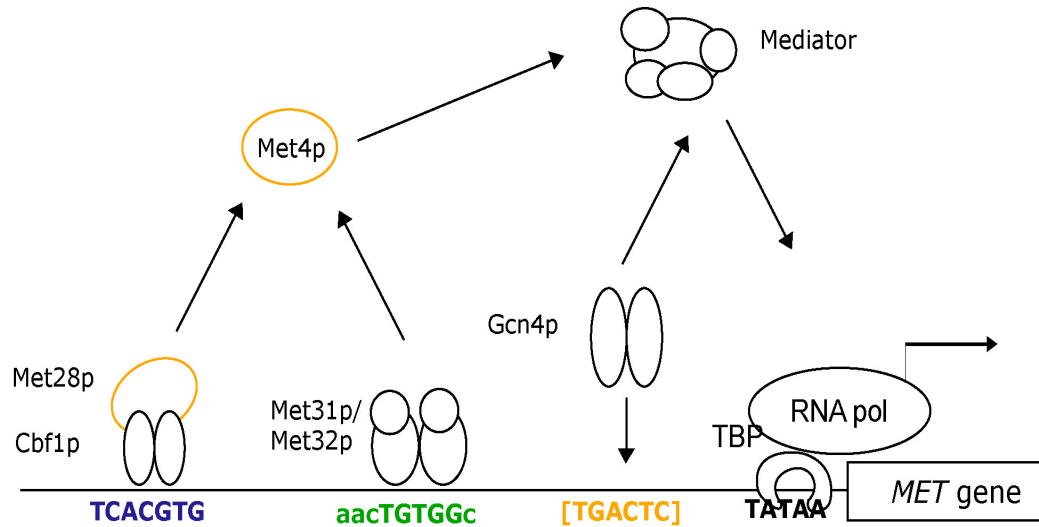
Gasch *et al.* (2000)

Hughes *et al.* (2000)

Natarajan *et al.* (2001)

Fauchon *et al.* (2002)

Promoter organization of yeast sulfur utilization genes

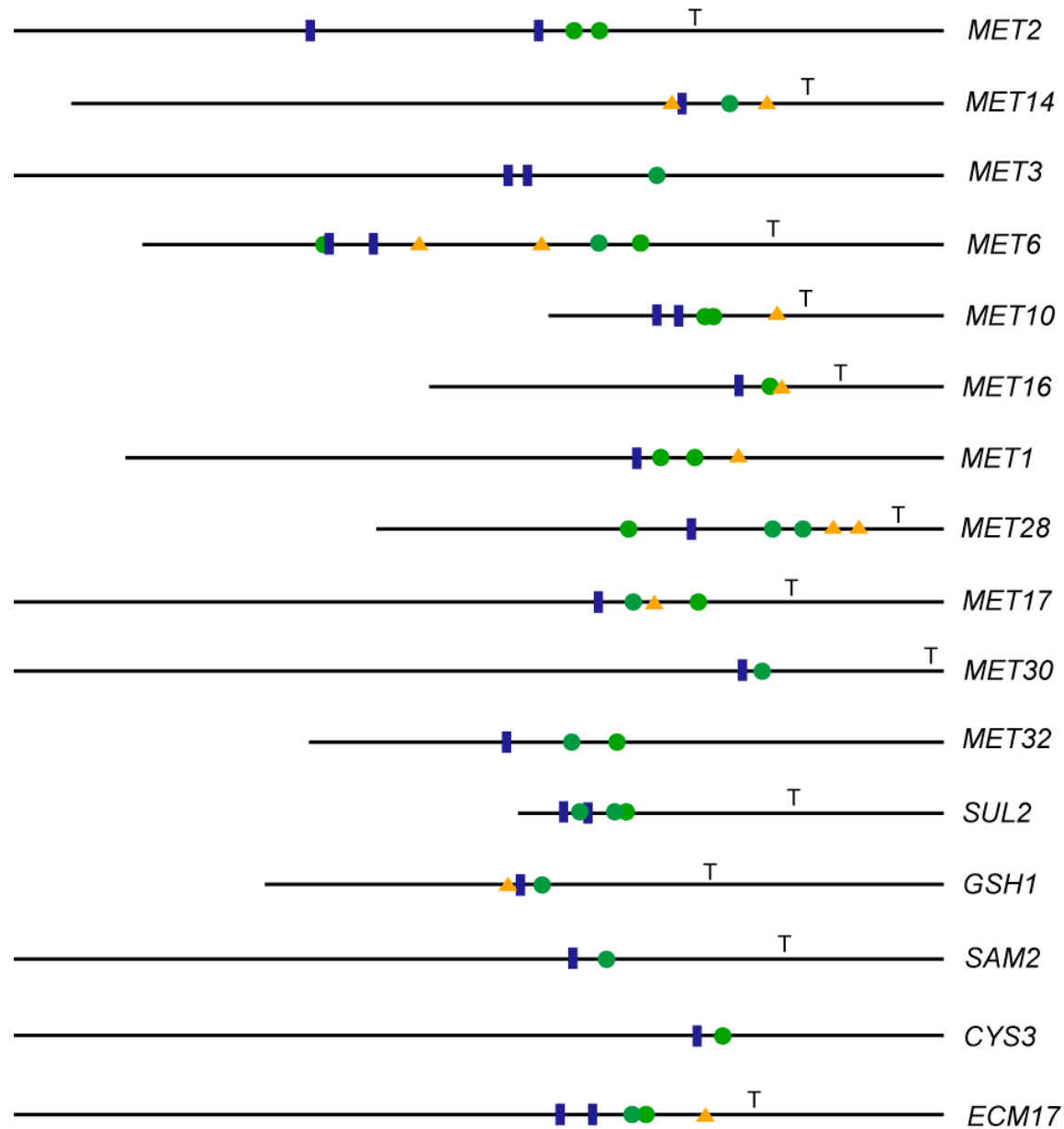


1) Genetic hypothesis testing

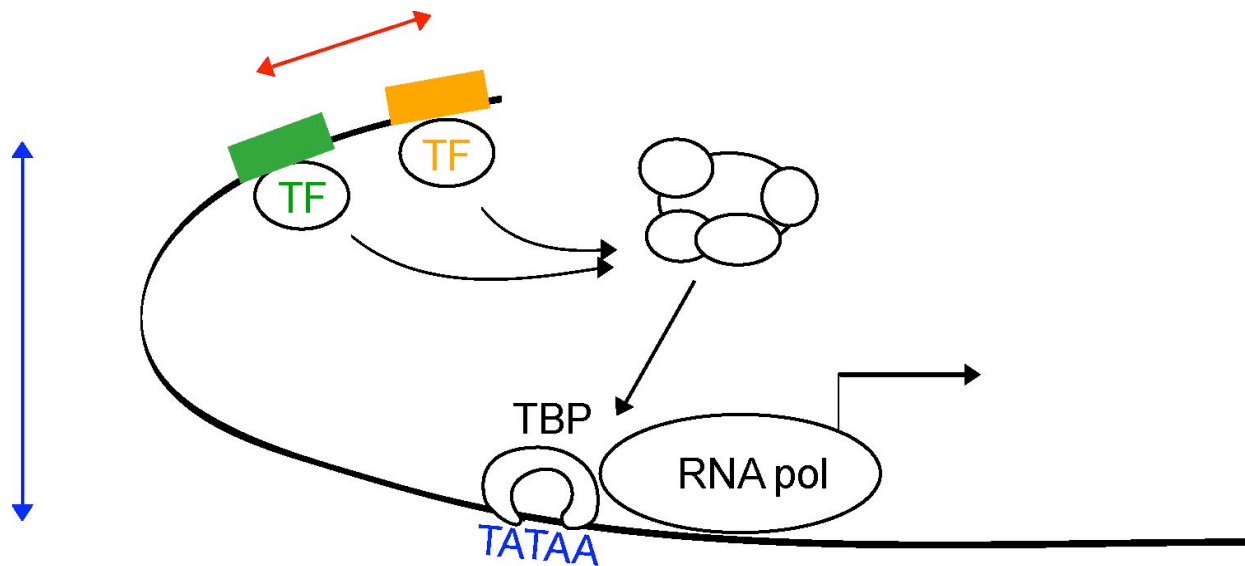
Predict Met4-regulated genes from microarray data

2) Define rules for promoter organization

Conserved binding sites in sulfur utilization promoters

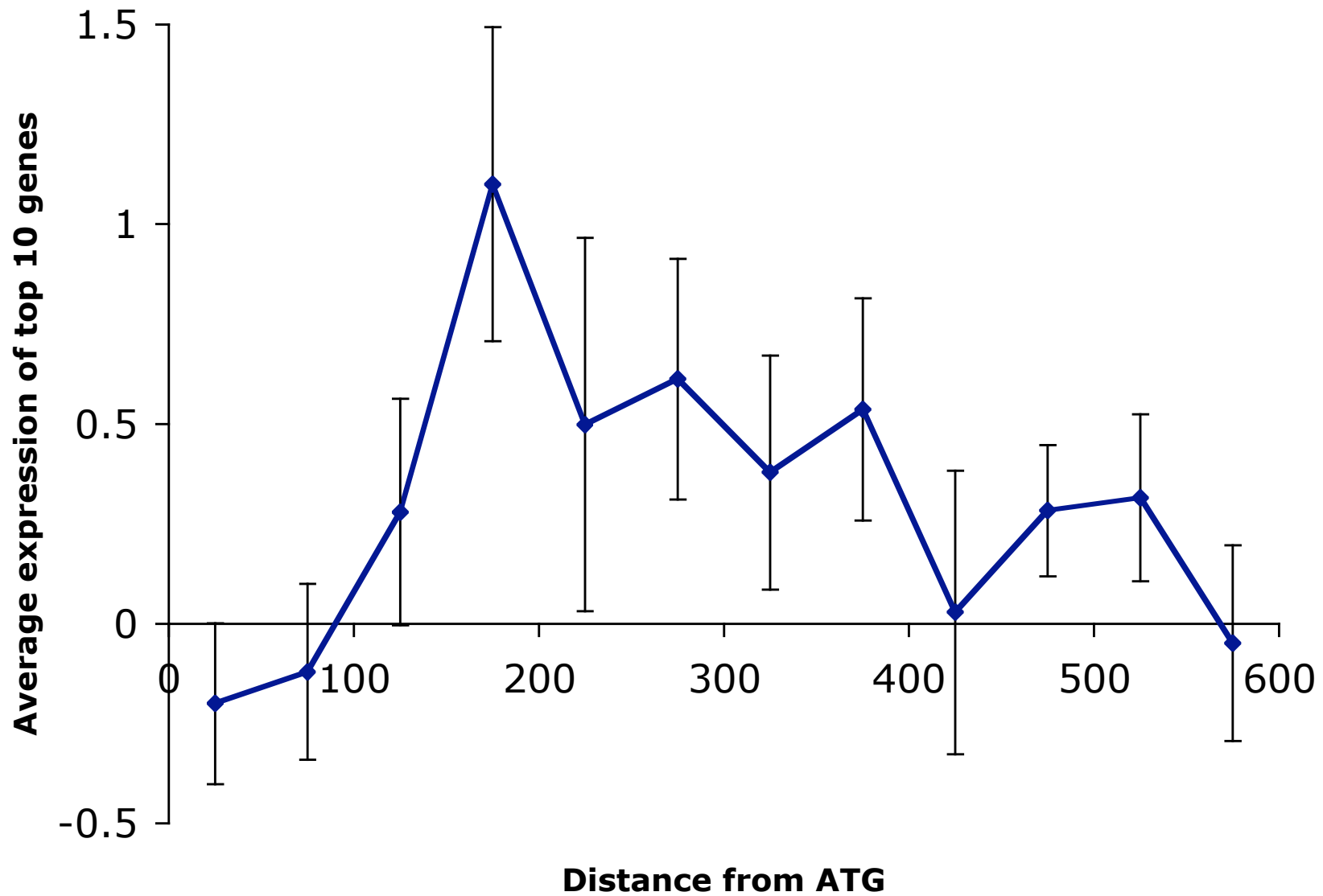


Sequence features of promoter organization

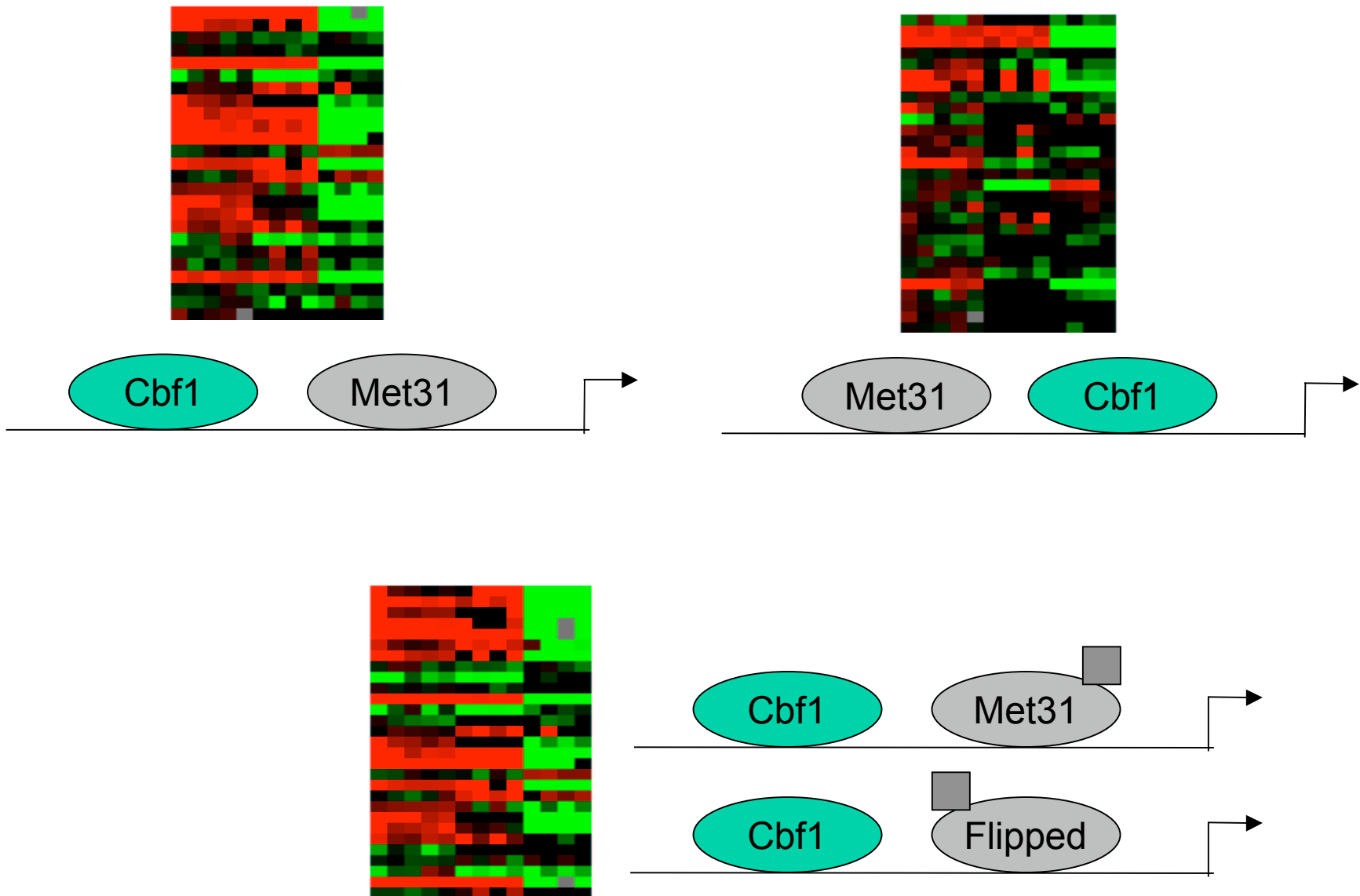


- 1) Distance to transcription start site
- 2) Orientation of transcription factor binding sites
- 3) Relative distance between binding sites

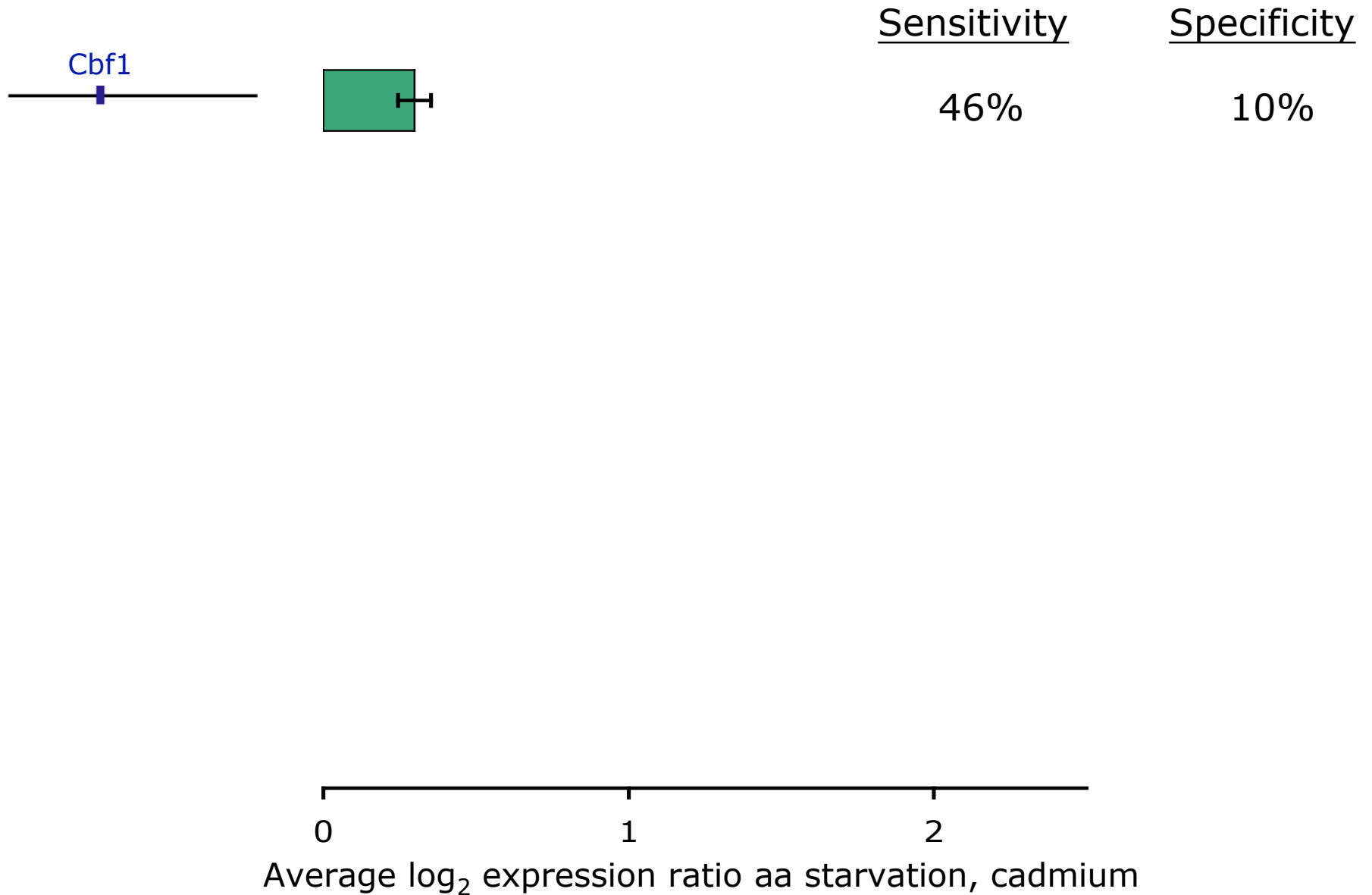
Promoter windows are associated with activation



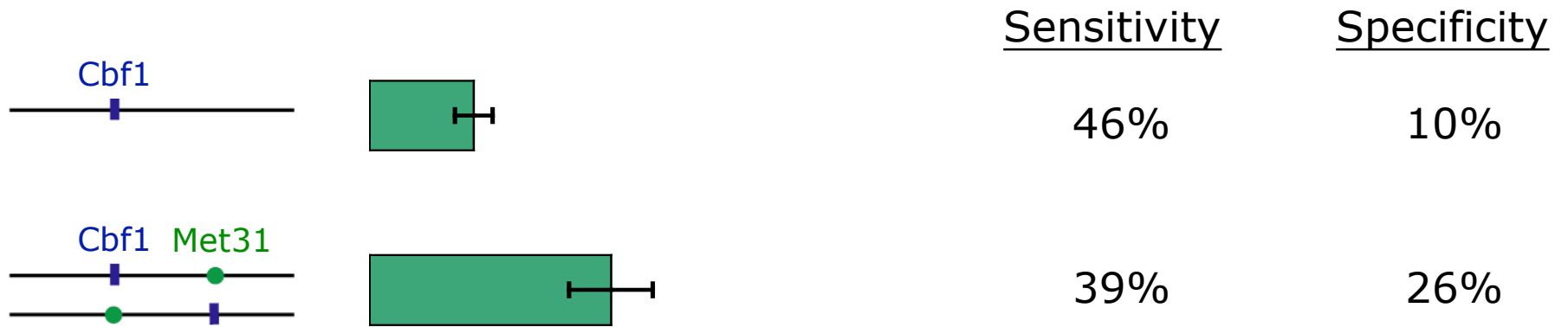
Order and orientation effects of binding sites



Promoter organization yields more specific predictions

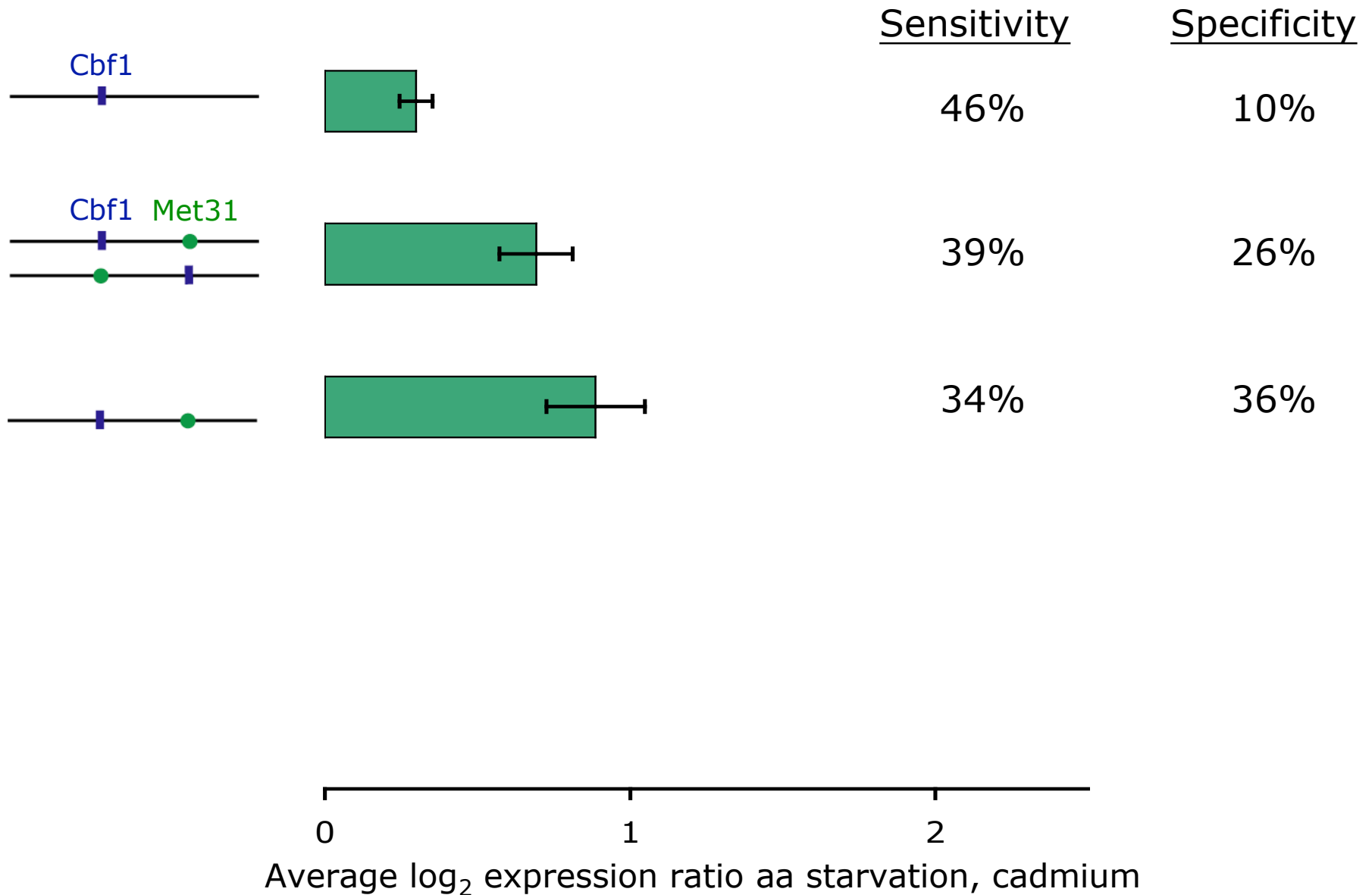


Promoter organization yields more specific predictions

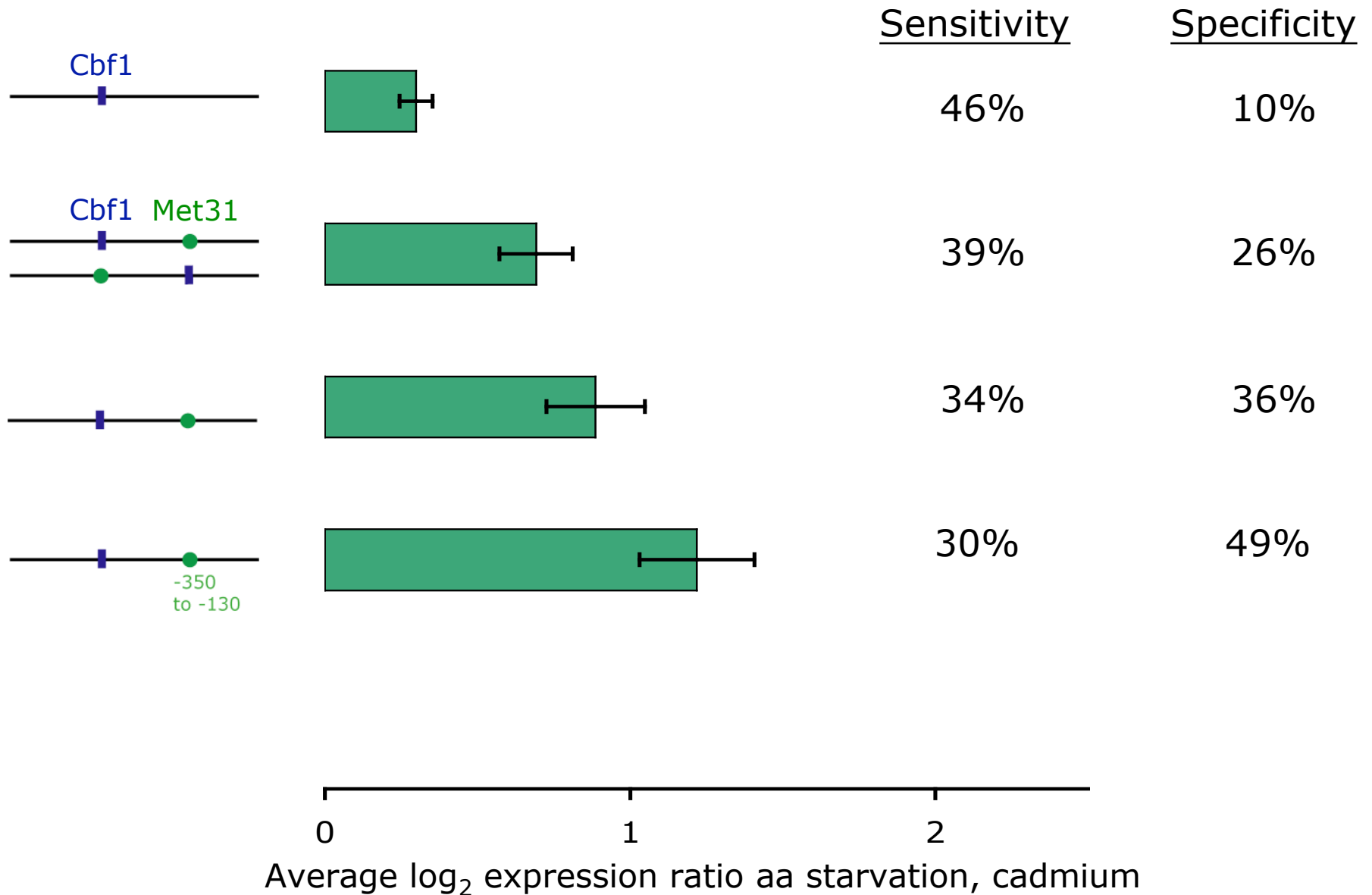


0 1 2
Average \log_2 expression ratio aa starvation, cadmium

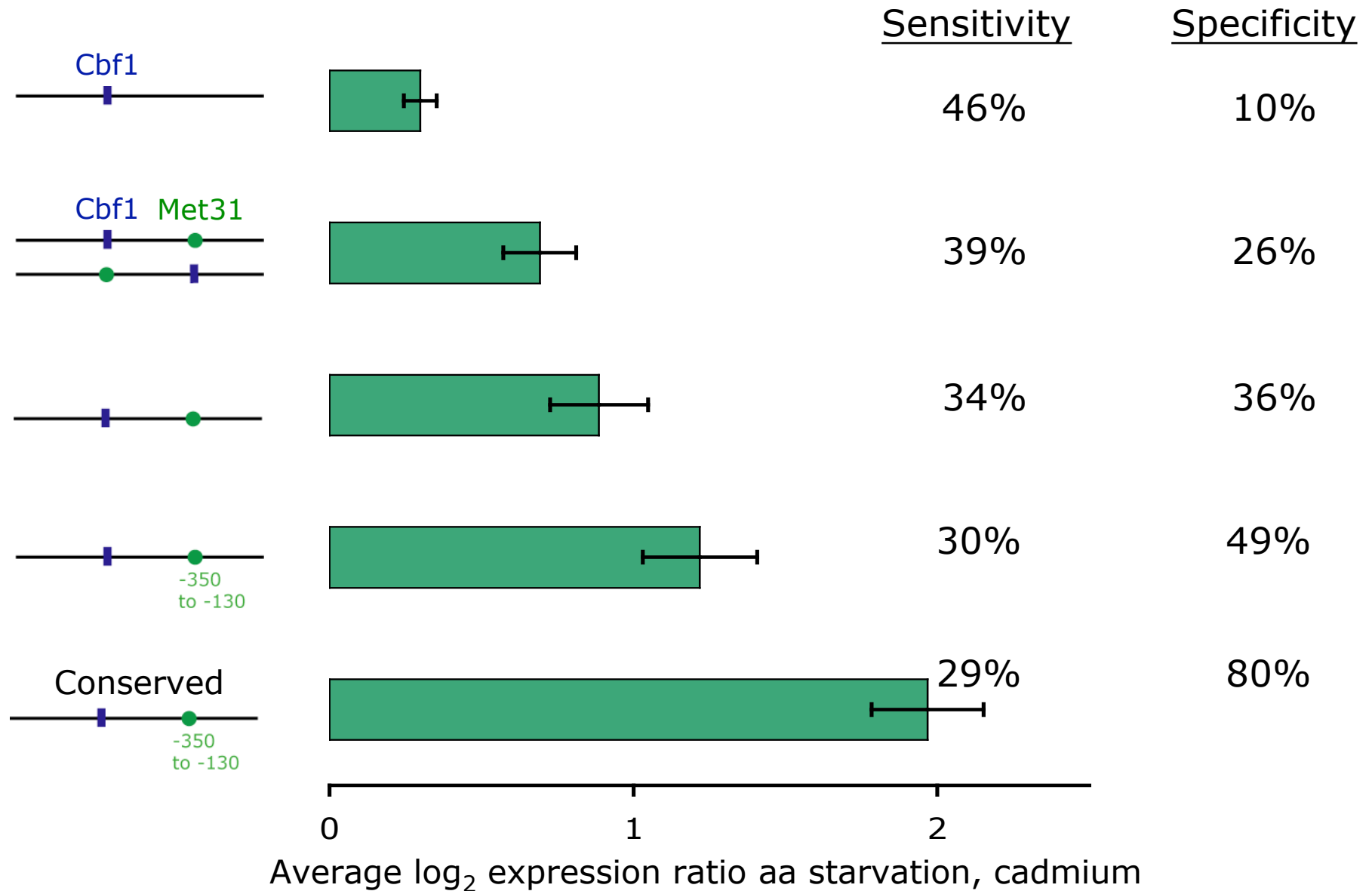
Promoter organization yields more specific predictions



Promoter organization yields more specific predictions



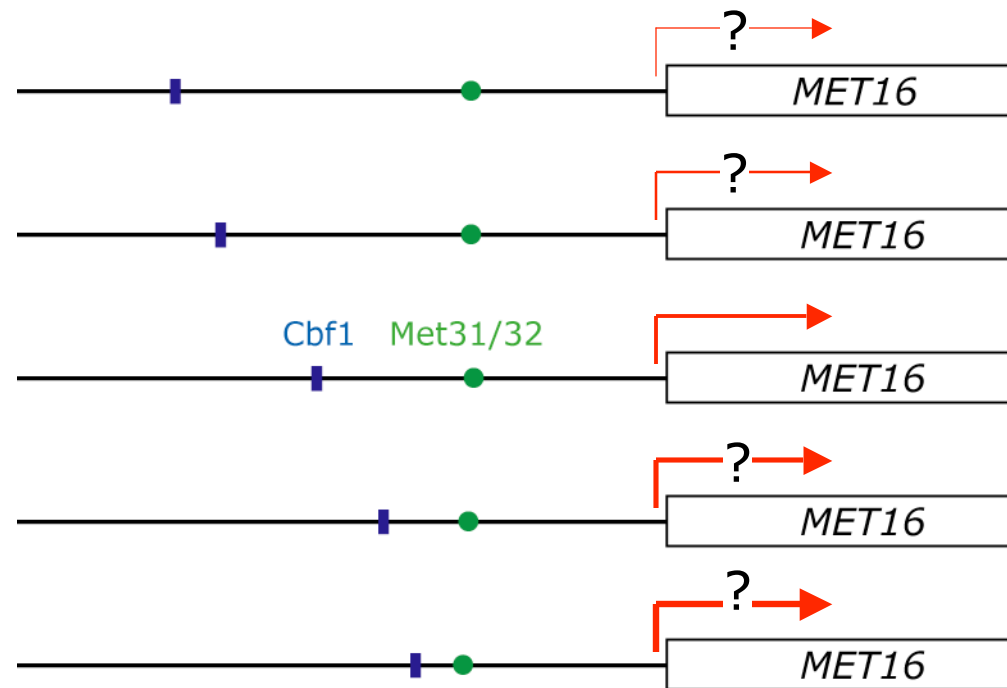
Promoter organization yields more specific predictions



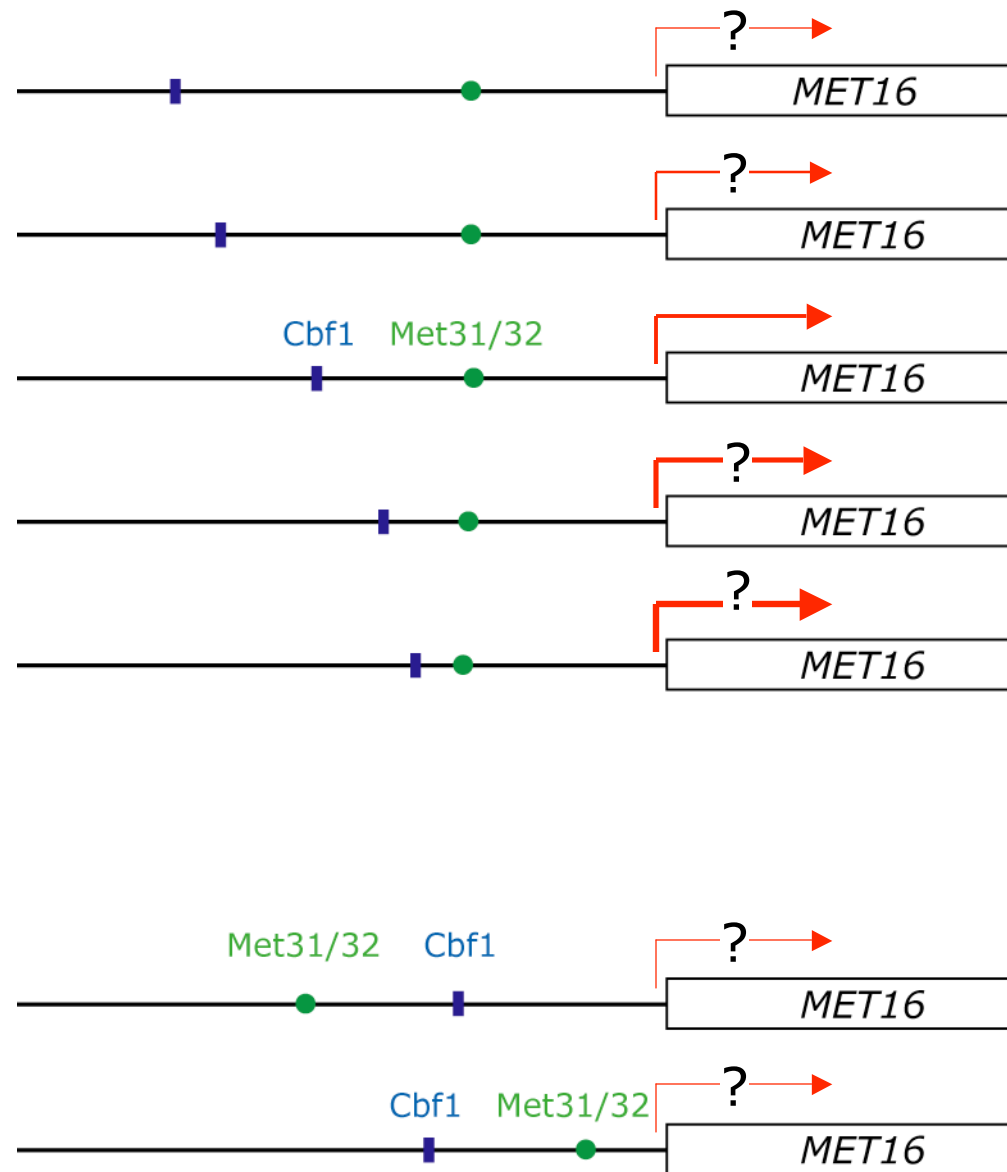
Experimental characterization of distance constraints



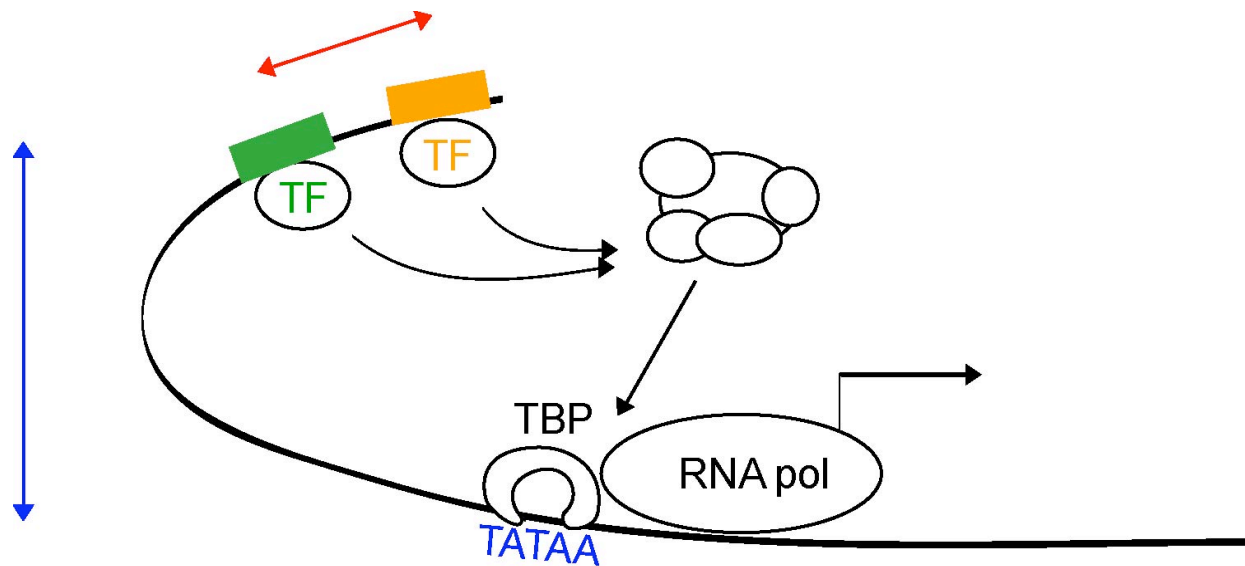
Experimental characterization of distance constraints



Experimental characterization of distance constraints



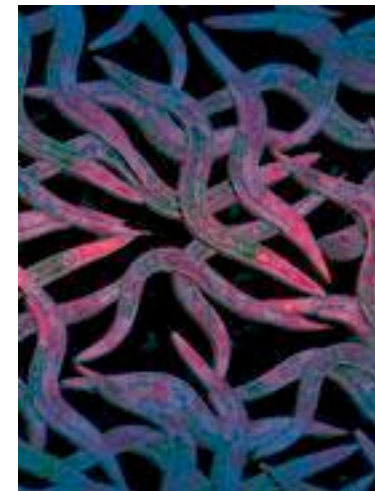
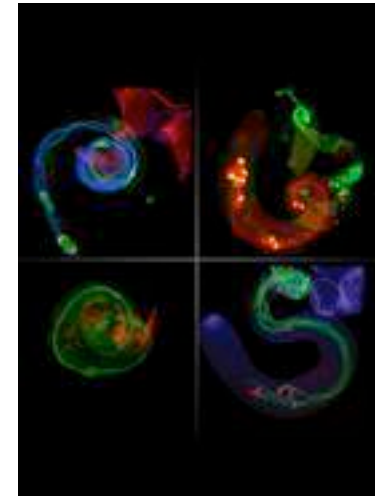
Summary



- Target genes can be accurately identified by genetic hypothesis testing of gene expression data
- Mechanisms for transcriptional regulation can be inferred **and tested** from characterization of promoter organization

Why is open access important?

- **Authors**
access to the largest possible audience
- **Readers**
access to the entire literature
- **Reuse of articles**
download, copy, print, archive
- **Full-text searching and mining**
beyond Boolean text searches



Acknowledgements

UC Berkeley

Alan Moses

Ryan Shultzaberger

David Nix

University of Toronto

Traci Lee

Paul Jorgensen

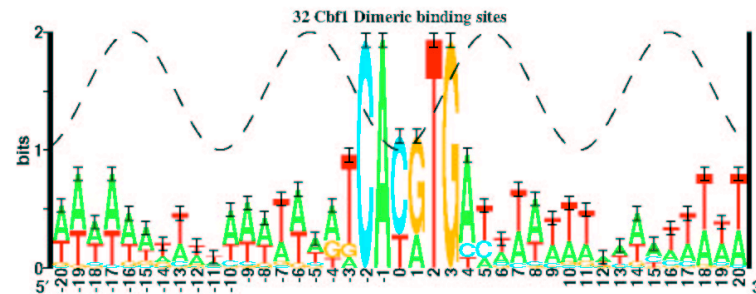
Mike Tyers

University of Wisconsin

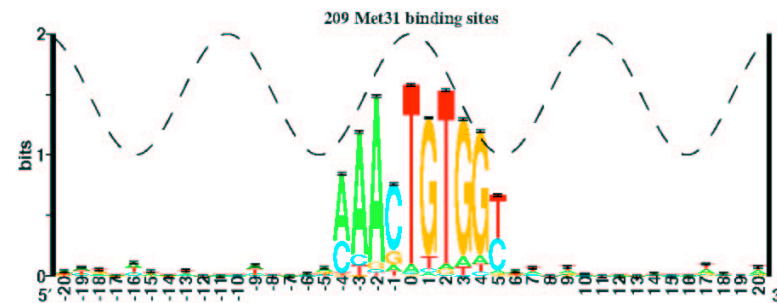
Audrey Gasch

Position weight matrices

Cbf1



Met31/32



Combining gene expression with Chromatin IP

